Comparative Analysis of Small Moleclues and Natural Plants Compounds as Therapeutic Inhibitors Targeting RdRp and Nucleocapsid Proteins of SARS CoV 2: An in Silico Approach

PAVAN KUMAR POLEBOYINA (pavanbiotech2012@gmail.com)
Osmania University Faculty of Science https://orcid.org/0000-0003-0393-3917

Research Article

Keywords: RdRp, N protein, SARS-CoV-2, COVID-19, Natural Plant Compounds, Small Molecules, Molegro Virtual Docker(RRID: SCR_000190)

Posted Date: June 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-655036/v1

Abstract

Nucleocapsid protein and RNA-dependent RNA polymerase (RdRp) activity in viral structural membrane, transcription and replication has been recognized as an attractive target to design novel antiviral strategies. The essential feature of the N protein of SARS COV 2 is to bind to the viral genome to promote the exact folding of the hammerhead ribozyme averting unproductive RNA confirmations and lead them to right into a helical capsid shape or RNP complex, whose packaging is crucial to viability. RdRp is an essential enzyme that helps in RNA synthesis by catalyzing the RNA template-dependent development of phosphodiester bonds. RdRp makes a complex with two cofactors nsp7 and nsp8 to play a key role in RNA synthesis, transcription and replication of the SARS-CoV-2. In our study we used small molecules and natural plants compounds as therapeutic inhibitors targeting RdRp and N proteins of SARS COV 2. Their structures were geometrically optimized and energetically minimized using Hyperchem software. Molecular docking was performed using Molegro virtual docker and top ligands were selected based on MolDock score,Rerank score and H-bonding energy. Our results showed that 9 compounds against N protein and 7 compounds against RdRp protein forming better inhibitory effect with most lowest MolDock score − 285.68kcal/mol and − 201.5kcal/mol respectively. we hope that these small molecules and natural plants compounds can inhibit the viral enzymes and helps the patients in reducing specific symptoms of SARS-CoV-2 infection. However in vivo experimental studies and clinical trials need to get that more favorable result.

Introduction

In Wuhan city of China, several pneumonia infected patients with unknown origin was identified on 31st December 2019[1]. The current pandemic SARS COV 2 reasons a public health emergency of international concern and significantly broken the world wide economy. COVID-19 is caused due to newly discovered and deadly coronavirus is a highly contagious disease, named severe acute respiratory syndrome coronavirus disease-2 (SARS Cov-2). The name ‘coronavirus’ is coined due to the crown-like projections on its surface as in Latin, ‘Corona’ means ‘halo’ or ‘crown’. On 13th January 2020, entire genome analysis becomes achieved found that novel corona virus (Gen Bank No. MN908947) officially called as SARS COV 2 previously known as SARS COV 1[2].

COV is a positive sense single stranded RNA(ssRNA) virus consists of 29,903 nucleotides and un translated sequences of 229 and 254 nucleotides on the 3’- and 5’- ends respectively. It is included in β corona virus genus, intently associated with the genomic organization of SARS COV identified in 2003[3]. The open reading frame(ORF) region of corona viruses contains specific genes which are encodes for spike, capsid and replicative proteins[4]. The two ORF’S(ORF1a, ORFlab) and polyproteins splits and encodes a structural and non-structural proteins. Structural proteins includes E protein (envelope), M protein (membrane), S protein (spike) and N (Nucleocapsid) proteins, while non-structural proteins are in a range from nsp1 to nsp16[5].

The replicase-transcriptase complex(RTC) consists of 16 nsps in which multiple enzymes, including nsp3 (papain like protease) ,3CL protease nsp5(chymotripsin like main protease), primase complex, RdRp(nsp12),nsp13(helicase)and nsp14(exoribonuclease)[6]. The nsp12 is the main content of the replication machinery which involves the transcriptional activity of RNA-dependent RNA polymerase (RdRp)[7]. It is a popular target for a selective antiviral strategy against SARS COV 2 because RNA synthesis by RNA-dependent RNA polymerase does not occur in mammalian cells[8]. The RdRP is present in all single stranded RNA viruses except in retroviruses. The Structural analyses of RNA viruses have shown that their RdRP protein contains of thumb, finger and palm domains, which are referred to as a closed, right-handed polymerase[9]. RdRp complex (nsp7-nsp8-nsp12) is the core component which is responsible for viral RNA replication and this complex also inclusive of nsp7-nsp8(nsp8-1) heterodimer and an additional subunit nsp8(nsp8-2) [10]. Inhibition of viral replication by targeting RdRp activity can becomes a powerful therapeutic approach[11].

SARS COV 2 N protein is firmly associated and interacts with viral genomic RNA structure and function. It contains phosphorylation sites at unstructured regions and it plays a major role in assembly and transcription of virus while functioning as an RNA chaperone[12,13]. Nucleocapsid protein was composed of alpha helix (21.24%), beta fold (16.71%), beta turn (6.92%), and random coil (55.13%). It has two main domains i.e. the RNA binding domain –N terminal domain (NTD)and the C-terminal dimerization domain (CTD). The NTD domain plays an important role in viral replication and transcription by binding to the 3’ end of the viral RNA genome due to positive amino acids electrostatic interactions. The major role of the N protein is to bind to the viral RNA to promote the exact folding of the hammerhead ribozyme avoiding unproductive RNA conformations. It makes them into a helical capsid structure or ribonucleoprotein (RNP) complex. Because of this function, it is suggested that the corona virus N protein is an RNA chaperone and performs a main role in viral genomic RNA replication[14,15]. Our present study has been undertaken to compare binding affinity analysis of various natural plants compounds and small molecules against RdRp and Nucleocapsid Proteins of SARS COV 2 by using Molecular Docking Methods[16].

Materials And Methods

Our present study involves retrieval of the 3d structure of target proteins and small molecules, natural plants compounds from Protein Data Bank[17] and Pubchem database[18] respectively. The proposed ligands designed and optimized in Hyperchem8[19]. Molecular docking experiments were performed by Molegro Virtual Docker(MVD2019_7_0_0)[20] and analyzed its data by Molegro Data modeler[21].

Retrieval of target proteins

3d structures of RdRp (RNA-dependent RNA Polymerase) Cryo-EM structure of the apo nsp12-nsp7-nsp8 complex protein (PDB ID 7BV1) and Crystal structure of SARS-COV-2 nucleocapsid protein N-terminal RNA binding domain (PDB ID 6M3M) was retrieved from Protein Data Bank. Before docking the water molecules, unwanted hetero atoms and other ligand compounds were removed by MVD.

Retrieval of Natural plants compounds and small molecules

A total 1,156 compounds were virtually screened and finally a total 56 hit lead optimized compounds were taken for consideration. The structures of Natural plants compounds and small molecules were retrieved from Pubchem and some structures were drawn in Hyperchem8 package; they were pre-optimized using the molecular mechanics force field (MM^+, AMBER) procedure. To obtain the conformers of ligands with lowest energy and semi-empirical method AM-1 was applied to the molecular structures. To avoid the local stable, each molecular structure was optimized several times with different starting points using the Polak-Rebiere algorithm, until the root-mean-square gradient is equal to 0.001 kcal $Å^{-1}$ mol^{-1}[22]. Then the energy minimizations of all ligand structures were done by using Hyperchem8. The first step was by calculating single point that used to determine the total molecular energy of the structure, the second step to determine the energy minimization algorithms that locate the flexible structure was using geometric optimization calculation(MM^+,AMBER force field)[23]. Then these total 56 structures were used for molecular docking calculations.

Prediction of drug binding cavities

The drug binding cavities in RdRp and Nucleocapsid proteins of SARS-COV-2 for Natural plants compounds and small molecule ligands are not well characterized. The amino acid residues responsible for cavity formation in RdRp and Nucleocapsid proteins were identified through MVD cavity detection algorithm. The program generally identified 5 different cavities. MVD identified 5 different cavities in both RdRp and N proteins.

Protein and ligand preparation

Protein

RdRp (RNA-dependent RNA Polymerase) Cryo-EM structure of the apo nsp12-nsp7-nsp8 complex protein (PDB ID 7BV1) and Crystal structure of SARS-COV-2 nucleocapsid protein N-terminal RNA binding domain (PDB ID 6M3M) were selected for the molecular docking studies using MVD. By Using the utilities provided in Molegro Virtual Docker all necessary H atom addition, valency checks, and protein preparation (protonation) were done and repair and rebuilt it. For precise docking it is important that the imported structures have been prepared accurately, that is the atom connectivity and bond orders are correct and partial atomic charges are assigned. PDB file often have poor or missing assignment of explicit hydrogen's, and the PDB file format cannot accommodate bond order information. Then the repair and rebuilt protein was saved in *.Mol format. The final structure was visualized and analyzed with SPDBV 4.10[24].

Ligand

The selected compounds were downloaded from Pubchem, zinc database and drug bank. Some torsion and peptide bond containing ligands were drawn and minimized its energy by using Hyperchem 8 software and imported to MVD workspace in *.Mol format. Before import the small molecules and all natural plants compounds undergoes series of steps that generates variation and optimization of the structure.

Molecular docking studies

Molegro Virtual Docker (MVD2019_7_0_0) program was used for validation of molecular docking. By comparing, the accuracy of MVD is higher than the other software like Glide, Surflex and FlexX. A total of 56 compounds were tested against RdRp and N proteins of SARS-COV-2. The MolDock scoring function was also set with a grid resolution of 0.30 Å. It was set at a maximum iteration of 1,500 with a simplex evolution size of 50 and a minimum of 10 runs were performed for each compound with threshold energy of 100. Additionally, the simplex evolution was set for 300 steps with a neighbor distance factor of 1.00. In MVD MolDock scoring function works based on a piecewise linear potential(PLP) and it is proposed by Gehlhaar et al yang et al.

The MolDock Score docking scoring function, EMolDock Score, is defined by the following energy terms Equation 1:

$$E_{MolDockscore} = E_{inter} + E_{intra}$$

Where, E_{inter} represents the ligand-protein interaction energy…Equation 2:

$$E_{inter} = \Sigma_{i \in ligand} \Sigma_{j \in protein} \left[E_{PLP}(r\ ij) + 332.0 \frac{qi\ qj}{4r2\ ij}\right]$$

Where E_{PLP} is the piecewise linear potential, the numerical value of 332.0 fixes the units of the electrostatic energy in kcal/mol. The second term describes the electrostatic interactions between charged atoms, and the internal energy of the ligand (E_{intra}) expressed by equation 3…

$$E_{intra} = \Sigma_{i \in ligand} \Sigma_{j \in protein} E_{PLR}(r\ ij) + \Sigma_{flexiblebonds} A[1-\cos(m.\theta-\theta_0)] + E_{clash}$$

The double summation contains all atom pairs in the ligand except those which are connected with two bonds or less. Second term is a torsional energy where θ is the torsional angle of the bond. E_{clash} assigns penalty of 1000 provided that the distance two heavy atoms is less than 2.0 Å[25]. The best pose of each compound was selected for subsequent ligand-protein interaction energy analysis[26]. Hydrogen bond interaction and its binding energy were observed between the amino acid residues on the target site with the functional group of small molecules and plants compounds.

ADMET properties

Top scoring molecules of N protein(salvianolic acid A, polyphyllin I, paeoniflorin, naringin, baicalin, ginsenoside rb I, alpha solanine, betanin and cairicoside I and RdRp protein(gentiopicroside, ginsenoside rb I, betanin, alpha solanine, naringin and polyphyllin I) were further used to estimate pharmacokinetic properties, drug-likeness and toxicity using the pkCSM[27], SwissADME tool[28] and molinspiration chemiinformatics tools[29].

Results And Discussion

Target protein conformation

The RdRp (RNA-dependent RNA Polymerase) Cryo-EM structure of the apo nsp12-nsp7-nsp8 complex protein (PDB ID 7BV1) with resolution 2.80Å and the Crystal structure of SARS-COV-2 nucleocapsid protein N-terminal RNA binding domain (PDB ID 6M3M) with resolution 2.70Å were retrieved from Protein Data Bank (PDB) database. The retrieved proteins checked structural conformations by ERRAT[30], PROCHEK[31] and PDBsum[32]. The validated 3d, secondary structural analysis and Ramachandran plot were shown in Fig.1, Fig.2.

The cavities (active sites) of the proteins RdRp and Nucleocapsid were obtained through MVD cavity detection algorithm and its place with volume illustrated in Fig.3.

Determination of ligand structures

The 56 structures of small molecules and natural plants compounds were used to find out their potential binding with the target proteins RdRp and N(Nucleocapsid). The compounds structural information and structural flexibility after single point energy and after geometrical optimization energy were shown in Table 1.

Validation of Docking Results

During the docking study, each compound is selected as best pose to determine the MolDock score, Rerank score, interactions energy, torsion angle and H-bonding against N and RdRp protein. The MolDock score, Rerank score, the total interaction energy between the pose and the target molecule, torsion angle and H-bonding energy of 56 compounds against N protein and RdRp protein are represented in Table 2 and Table 3 respectively. In our study we compared the binding efficacy of ligands against protein with and without energy minimization. We found that energy minimized compounds shows better binding results[33]. Among all docked 56 ligands, 9 compounds shows better binding energy with N protein and 7 compounds with RdRp protein. Among all these 56 compounds 9 and 7 compounds shows better inhibitory effect on N protein and RdRp protein respectively.

N protein results

Out of all 56 compounds, Cairicoside I, Ginsenoside rb1, polyphyllin I, Gambogic acid, Betanin, and Alpha solanine revealed the most lowest MolDock score on N protein which is -285.68kcal/mol, -172.65kcal/mol, -166.78kcal/mol, -164.94kcal/mol, -160.78kcal/mol, -160.01kcal/mol respectively and Naringin, Salvianolic acid A, Betanin, Paeoniflorin, Ginsenoside rb1, Polyphyllin I, Baicalin shows excellent H-Bonding on N protein which is -23.82 kcal/mol, -23.26 kcal/mol, -19.80 kcal/mol, -19.32 kcal/mol, -18.50kcal/mol, -18.28kcal/mol -18.27kcal/mol respectively. By comparing entire parameters out of 56, 9 compounds were selected for better inhibitory studies, which are Alpha solanine binds into the active site of N protein with MolDock score -160.01kcal/mol & binding site consists of amino acid residues like Asp129A, Lys62B, Glu68B, Lys128A, Asp64B, Arg89A, Glu119A, Pro118A, Asn154D, Ile131B, Ile132B, Trp132B, Trp133B, Ala126B, Gly125B, Asn127B, His146D, Ile147D, Trp53D, Asn78D, Asn151A, Asn49A, Asn155D, Thr50A. It forms hydrogen bonds on Arg89A, Asn127B, Asn155D, Gly125B, Asn78D, Asn49A, Ile131B amino acid residues with binding energy -17.11kcal/mol. Baicalin binds into the active site of N protein with MolDock score -105.37kcal/mol & binding site consists of amino acid residues like Trp109B, Lys66B, Lue65B, Asp64B, Ile132B, Arg90B, Gly130B, Ile131B, Trp133B, Asn49A, Asp129B,Lys128B, Asn154B, Asn151D, Asn127B, Asn155B, Arg150D, Trp53B, Thr149D. it forms hydrogen bonds on Asn151D, Thr149D, Asn127B, Asn155D, Asn154D, Gly130B, Asp129B, Asp64B With binding energy -18.27kcal/mol. Betanin binds into the active site of N protein with MolDock score -160.78kcal/mol & binding site consists of amino acid residues like Trp133B, Lys128B, Asn127B, Trp53D, Asn78D, Asn49A, Thr50A, Ala51A, Arg89A, Ala91A, Arg90A, Thr92A, Lys66B, Glu63B, Pro169B, Lys170B. it forms hydrogen bonds on Asn127B, Arg89A, Arg90A, Thr92A, Glu63B, Lys66B With binding energy -19.8kcal/mol. Cairicoside I binds into the active site of N protein with MolDock score -285.68kcal/mol & binding site consists of amino acid residues like Trp53D, Ile147D, Ile158D, Asn78D, Asn155D, Asn154D, Asn127B, Asn151A, Thr50A, Pro118A, Tyr112A, Ser52A, Ala51A, Gly125B, Ile131B, Trp133B, Ile132B, Arp69B, The67B, Pro68B, Val159C, Tyr110A, Asp64B, Trp109B, Lue65B, The67B. it forms hydrogen bonds on Asn49A, Asn127B, Asn154D, Asn78D, Thr50A, Pro68B, Lys66B With binding energy -16.8kcal/mol. Ginsenoside rb1 binds into the active site of N protein with MolDock score -172.65kcal/mol & binding site consists of amino acid residues like Arg108A, Arg93A, Thr92A, Ana91A, Tyr110A, Ser52A, Tyr112A, Ala51A, Arg89A, Thr50A, Tro118A, Asn49A, Lys66B, Asn155D, Asn154D, Asn151D, Asn127B, Gly125B, Trp133B, Ile132B, Ile131B, Lys128B, Ala126B, Thr149D, Asn151D, Arg150D, Trp53D. it forms hydrogen bonds on Asn155D, Asn151D, Thr149D, Asn127B, Asn49A, Lys128B, Ile131B, Thr50A, Tyr112A With binding energy -18.50kcal/mol. Naringin binds into the active site of N protein with MolDock score -145.45kcal/mol & binding site consists of amino acid residues like Arg90A, Arg89A, Thr92A, Tyr110A, Tyr112A, Ser52A, Pro118A, Glu63B, Ala51A, Thr50A, Asn49A, Lys66B, Trp153B, Ile132B, Phe67B. it forms hydrogen bonds on Trp133B, Thr50A, Thr92A, Glu63B, Lys66B, Arg89A, Ser52A, Tyr122A With binding energy -23.82kcal/mol. Paeoniflorin binds into the active site of N protein with MolDock score -114.76kcal/mol & binding site consists of amino acid residues like Trp53D, Asn155D, Thr50A, Asn49A, Asn127B, Asp129B, Lys128B, Ala126B, Gly125B, Gly130B, Tle131B, Trp133B, Lys66B. it forms hydrogen bonds on Asn127B,Thr50A, Ile131B, Ala126B, Lys128B, ASN 49A With binding energy -19.32kcal/mol. Polyphyllin I

binds into the active site of N protein with MolDock score -166.78kcal/mol & binding site consists of amino acid residues like Thr50D, Thr149D, Gly148D, Tle147D, Trp53D, Asn51D, Asn155D, Asn154D, Asp129D, lys128D, Asn127B, gly130B, Thr50A, Ala51A, Lys66B, Pro152A, Asn49A, Ile132B, Trp133B, Phe67B, Pro68B, Arg69B, Gln161C. it forms hydrogen bonds on Arg69B, Trp133B, Phe67B With binding energy -18.28kcal/mol. Salvianolic acid A binds into the active site of N protein with MolDock score -143.71kcal/mol & binding site consists of amino acid residues like Tyr112A, Arg89A, Pro118A, Thr50A, Asn154D, Asn155D, Trp53D, Asn127B, Asn49A, Lys66B, Asp64B, Lys128B, Asp129B, Ile131B, Gly130B, Arg90B. It forms hydrogen bonds on Asn49A, Thr50A, Arg89A, Gly130B, Asp129B, Ile131B, Lys128B, Asn127B, and Tyr112A With binding energy -23.26kcal/mol. Fig.4 illustrates hydrogen bonding with amino acids of corona virus protein N.

Rdrp Protein Results

Out of all 56 compounds, Alpha solanine, Betanin, Cairicoside I, Gambogic acid, Ginsenoside rb I, Salvianolic acid A and Lycopene revealed the most lowest MolDock score on RdRp protein which is -151.07kcal.mol, -156.39kcal/mol, -201.55kcal/mol, -164.15kcal/mol, -147.04kcal/mol, -150.83kcal/mol, -159.57kcal/mol respectively and Polyphyllin I, Naringin, Ginsenoside rb 1, Gentiopicroside, Betanin and Alpha solanine shows excellent H-Bonding on RdRp protein which is -18.45kcal/mol, -18.25kcal/mol, -19.13kcal/mol, -20.08kcal/mol, -17.11kcal/mol and -16.47kcal/mol respectively. By comparing entire parameters out of 56 compounds, 7 compounds were selected for better inhibitory studies, which are Alpha solanine binds into the active site of RdRp protein with MolDock score -151.07kcal/mol & binding site consists of amino acid residues like Val557A, Ser682A, Lys545A, Thr556A, Arg555A, Asp62A, Asp760A, Asp452A, Arg624A, Cys622A, Arg553A, Tyr455A, Lys551A, Lys621A, Pro620A, Lys798A, Tyr619A and Ala554A. It forms hydrogen bonds on Lys621A, Cys622A, Tyr619A, Asp760A, and Asp623A amino acid residues with binding energy -16.47kcal/mol. Betanin binds into the active site of RdRp protein with MolDock score -156.39kcal/mol & binding site consists of amino acid residues like Lys545A, Lys500A, Val557A, Gly683A, Thr556A, Ser685A, Ala554A, Asp452A, Arg553A, Tyr455A, Lys621A, Arg624A, Asp623A, Asn691A, Ser759A, Thr687A, Ala688A, Ser759A and Lys545A. It forms hydrogen bonds on Ala688A, Thr556A, Arg624A, Arg555A, Asp452A, Lys621A, Asp623A and Lys545A amino acid residues with binding energy -17.11kcal/mol. Gentiopicroside binds into the active site of RdRp protein with MolDock score -89.50kcal/mol & binding site consists of amino acid residues like Tyr619A, Cys622A, Pro620A, Lys621A, Tyr458A, Asp623A, Arg624A, Tyr455A, Tyr458A, Arg553A, Asp452A, Ala554A, Arg555A and Tyr556A. It forms hydrogen bonds on Tyr619A, Cys622A, Asp623A, Lys621A, Arg555A and Ala554A amino acid residues with binding energy -20.08kcal/mol. Ginsenoside rb1 binds into the active site of RdRp protein with MolDock score -147.04kcal/mol & binding site consists of amino acid residues like Ala688A, Thr687A, Ser759A, Asn691A, Asp760A, Ser682A, Asp623A, Tyr619A, Cys622A, Arg555A, Thr556A, Asp618A, Pro620A, Lys621A, Asp452A, Ala554A, Tyr455A, Lys798A and Arg553A. It forms hydrogen bonds on Lys798A, Asp618A, Arg555A, Asp452A, Ala554A, Lys621A, Arg553A, Asp760A and Ser759A amino acid residues with binding energy -19.13kcal/mol. Naringin binds into the active site of RdRp protein with MolDock score -134.00kcal/mol & binding site consists of amino acid residues like Ala550A, Ser549A, Lys551A, Ala554A, Arg555A, Arg553A, Asp452A, Thr556A, Tyr455A, Arg624A, Lys621A, Ser682A, Asp623A, Cys622A, Thr687A and Asn691A. It forms hydrogen bonds on ASN 691A, Arg624 A, Thr556A, Lys621A, Cys622A and Asp623A amino acid residues with binding energy -18.25kcal/mol. Polyphyllin I binds into the active site of RdRp protein with MolDock score -140.21kcal/mol & binding site consists of amino acid residues like Thr556A, Ala554A, Asp452A, Ser682A, Thr687A, Arg553A, Tyr455A, Arg624A, Asp623A, Ser759A, Leu758A, Asp760A, Cys622A, Lys798A, Pro620A, Asp648A, Tyr619A, Lys798A and Lys551A. It forms hydrogen bonds on Asp623A, Thr556A, Arg555A, Ala554A, Asp452A, Ser759A and Lys798A amino acid residues with binding energy -18.45kcal/mol. Cairicoside I binds into the active site of RdRp protein with MolDock score -201.55kcal/mol & binding site consists of amino acid residues like Lys500A, Ala685A, Asp684A, Gly683A, Ala558A, Val557A, Ser682A, Thr687A, Ala688A, Ser759A, Leu758A, Asp761A, Cys813A, Ser814A, Asp760A, Tyr619A, Lys545A, Asp623A, Cys622A, Thr556A, Arg624A, Asp452A, Tyr455A, Arg553A, Lys621A and Pro620A. It forms hydrogen bonds on Asp684A, Asp761A, Asp760A, Lys545A, Lys621A, Arg624A and Ser628A amino acid residues with binding energy -7.29kcal/mol. Fig.5 illustrates hydrogen bonding with amino acids of corona virus protein RdRp.

Fig.6 and Fig.7 shows, steric interactions with residues of COVID-19 proteins N and RdRp respectively. Fig.8 and Fig.9 represents Ligand docked against the crystal protein structures of COVID-19 Nucleocapsid and RdRp Proteins respectively. The amino acid residues around active site and Docked against N and RdRp protein were illustrate in Table 4 and Table 5 with structures at Fig.10 and Fig.11 respectively.

ADMET results

In the docking study, more negative BE corresponded to the strong binding of selected compounds to the target proteins. It is a fact that weaker binding will ultimately have a rapid dissociation rate[34]. In this study, the selected compounds exhibited lower binding energy which means strong binding with the N and RdRp proteins, suggesting that these compounds could be utilize as an inhibitors of the RdRp and Nucleocapsid proteins to combat COVID-19. According to the Rule of Five, a molecule might be no longer orally active if it violates or greater of the 4 rules. But the rules are not suitable for natural products because they are complicated and they have been deduced from relatively simple small molecules[35]. However, we have filtered the compounds based on the binding efficacy on Both Proteins of COVID-19. The results of the pkCSM, SwissADME tool and molinspiration chemiinformatics tools shows that top scored compounds pharmacokinetic properties (N and RdRp protein inhibitory compounds) (Table 6).

Discussion

Developing drugs to decrease the symptoms of SARS-Cov-2 infection is the main task now in worldwide. Now a days new research is published on SARS-CoV-2 and proposes a new drug among drug repurposing medicines for the possible treatment of COVID-19. In our study, we targeted two essential proteins(N and RdRp) for identifying small molecules and natural plant compounds as therapeutic agents, using in-silico analysis. The three dimensional proteins structures of N and RdRp were retrieved from RCSB Protein Data Bank. According to our results, some natural plant compounds

were detected against N protein including Alpha solanine, Baicalin, Betanin, Cairicoside I, Ginsenoside rb1, Naringin, Paeoniflorin, Polyphyllin I, Salvianolic acid A; among them, Alpha solanine, Betanin, Cairicoside I, Ginsenoside rb1, Naringin, Polyphyllin I were identified to have interaction with RdRp too. This study has identified 56 natural plant compounds, some small molecules with higher binding affinity and interaction with N and RdRp proteins active pocket residues. The result of this and other similar in-silico studies on Natural plant compounds and small molecules are promising for further in vitro, in vivo and clinical trial studies of SARS-CoV-2 treatment.

Conclusion

The present study involves the capability Natural Plant compounds and small molecules from medicinal plants in forming stable interactions with N and RdRp proteins of the SARS-CoV-2. Developing of herbal drugs with minimum side effects is the better opportunity to explore the medicinal and other biological properties of Natural plant compounds. To identify its inhibitory effects and usefulness, it is mandatory to focus on visualization and identification of unused Natural Plant Compounds in a particular disease over the world. Then it's far emphasized on extraction,isolation and characterization of Natural Plant Compounds and its small molecules is a gift of nature in a rational and scientific way. In this study, two main COVID-19 viral proteins, N and RdRp were selected to dock against Natural Plant Compounds and small molecules. MolDock score, Rerank score, H-Bonding interactions and free energy calculations suggest favorable binding of ligands such as Alpha solanine, Betanin, Cairicoside I, Ginsenoside rb1, Naringin and polyphyllinI. These compounds shows harmonious inhibitory effect on both N and RdRp proteins and several amino acid residues were observed to participate in such interactions. Overall, the findings in the article proposes the Natural Plant Compounds and small molecules as stable N and RdRp protein catalytic active site binders, synergistic with the experimentally known drug-N and RdRp interactions. However, many of them to be further investigate experimentally against COVID-19 in vitro and in vivo.

Declarations

Data availability

Data supporting the productivity of this investigation are available from the corresponding author upon request.

Conflict of interests

The authors declare that there is no conflict of interest. No additional benefits will be received from a third party directly or indirectly by the authors.

Author's contributions

SMITA C PAWAR conceptualized, designed, interpreted data and edited manuscript; PAVAN KUMAR POLEBOYINA conceptualized and designed the study and conducted the study. All authors have approved the manuscript in the current form.

Acknowledgments

Thankful CSIR-UGC for providing Fellowship to PAVAN KUMAR POLEBOYINA, reference Nos. 1621 and thankful to staff of Department of Genetics & Biotechnology, Osmania University for encouragement and support in Bioinformatics/software facilities. Thankful to Rene Thomsen, Molexo, Denmark for providing MVD trial version (RRID: SCR_000190).

Funding

The authors did not receive support from any organization for the submitted work.

Ethical approval

Not applicable

Consent to participate

Not applicable

Consent to publish

Not applicable

Tables

Please see the supplementary files section to view the tables.

References

1. Park, S.E. Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-COV-2; Coronavirus Disease-19). *Korean J. Pediatr.* 2020, *63*, 119–124.

2. In search of RdRp and Mpro inhibitors against SARS CoV-2: Molecular docking, molecular dynamic simulations and ADMET analysis | Elsevier Enhanced Reader Available online: https://reader.elsevier.com/reader/sd/pii/S0022286021006219?token=AAE1D637C7288A692C02D79789516BA80F2E8C23B0C7E62CFB8DD8D2B3EA37EC5095959C05AE949B3566B67C8630B6ED&originRegion=eu-west-1&originCreation=20210606193039 (accessed on Jun 7, 2021).

3. Inhibitory efficacy of RNA virus drugs against SARS-CoV-2 proteins: An extensive study | Elsevier Enhanced Reader Available online: https://reader.elsevier.com/reader/sd/pii/S0022286021002830?token=81295A7DAA10F91A2DB847833068D13246AC9D9C6491F87A776EBBA8E17DEE451F9B12BD35BD2CF84C4998143CF5E3FC&originRegion=eu-west-1&originCreation=20210606195144 (accessed on Jun 7, 2021).

4. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses | Elsevier Enhanced Reader Available online: https://reader.elsevier.com/reader/sd/pii/S2090123220300540?token=6DDBB062F469ED88128E0BEA7AF62A17089CB020930AB21F56295339E534532E592A1609BC5109784E9655941E42AC33&originRegion=eu-west-1&originCreation=20210606200221 (accessed on Jun 7, 2021).

5. Mahrosh, H.S.; Mustafa, G. An in silico approach to target RNA-dependent RNA polymerase of COVID-19 with naturally occurring phytochemicals. *Environ. Dev. Sustain.* **2021**, doi:10.1007/s10668-021-01373-5.

6. Zhu, W.; Chen, C.Z.; Gorshkov, K.; Xu, M.; Lo, D.C.; Zheng, W. RNA-Dependent RNA Polymerase as a Target for COVID-19 Drug Discovery. *SLAS Discov.* 2020, *25*, 1141–1151.

7. Biswal, M.; Diggs, S.; Xu, D.; Khudaverdyan, N.; Lu, J.; Fang, J.; Blaha, G.; Hai, R.; Song, J. Two conserved oligomer interfaces of NSP7 and NSP8 underpin the dynamic assembly of SARS-CoV-2 RdRP. *Nucleic Acids Res.* **2021**, *12*, 1–11, doi:10.1093/nar/gkab370.

8. Ramanathan, A.; Pullara, F.; Kumar, T.D.; Shaikh, N.; Kumar, U.S.; Priya Doss, G.C.; Zayed, H. Structure-Based Virtual Screening to Identify Novel Potential Compound as an Alternative to Remdesivir to Overcome the RdRp Protein Mutations in SARS-CoV-2. *Front. Mol. Biosci. | www.frontiersin.org* **2021**, *1*, 645216, doi:10.3389/fmolb.2021.645216.

9. Machitani, M.; Yasukawa, M.; Nakashima, J.; Furuichi, Y.; Masutomi, K. RNA-dependent RNA polymerase, RdRP, a promising therapeutic target for cancer and potentially COVID-19. *Cancer Sci.* 2020, *111*, 3976–3984.

10. Jiang, Y.; Yin, W.; Xu, H.E. Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information . **2020**.

11. Aftab, S.O.; Ghouri, M.Z.; Masood, M.U.; Haider, Z.; Khan, Z.; Ahmad, A.; Munawar, N. Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a potential therapeutic drug target using a computational approach. *J. Transl. Med.* **2020**, *18*, 1–15, doi:10.1186/s12967-020-02439-0.

12. Tugaeva, K. V.; Hawkins, D.E.D.P.; Smith, J.L.R.; Bayfield, O.W.; Ker, D.S.; Sysoev, A.A.; Klychnikov, O.I.; Antson, A.A.; Sluchanko, N.N. The Mechanism of SARS-CoV-2 Nucleocapsid Protein Recognition by the Human 14-3-3 Proteins:SARS-CoV-2 N association with host 14-3-3 proteins. *J. Mol. Biol.* **2021**, *433*, 166875, doi:10.1016/j.jmb.2021.166875.

13. Batra, M.; Tian, R.; Zhang, C.; Clarence, E.; Sacher, C.S.; Miranda, J.N.; De La Fuente, J.R.O.; Mathew, M.; Green, D.; Patel, S.; et al. Role of IgG against N-protein of SARS-CoV2 in COVID19 clinical outcomes. *Sci. Rep.* **2021**, *11*, 1–9, doi:10.1038/s41598-021-83108-0.

14. Gao, T.; Gao, Y.; Liu, X.; Nie, Z.; Sun, H.; Lin, K.; Peng, H.; Wang, S. Identification and functional analysis of the SARS-COV-2 nucleocapsid protein. *BMC Microbiol.* **2021**, *21*, 1–10, doi:10.1186/s12866-021-02107-3.

15. Tatar, G.; Ozyurt, E.; Turhan, K. Computational drug repurposing study of the RNA binding domain of SARS-CoV-2 nucleocapsid protein with antiviral agents. *Biotechnol. Prog.* **2021**, *37*, doi:10.1002/btpr.3110.

16. Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. Molecular docking: a powerful approach for structure-based drug discovery. Current computer-aided drug design. *Curr. Comput. Aided Drug Des.* **2011**, *7*, 146–157.

17. RCSB PDB: Homepage Available online: https://www.rcsb.org/ (accessed on May 24, 2021).

18. PubChem Available online: https://pubchem.ncbi.nlm.nih.gov/ (accessed on May 24, 2021).

19. Daoud, I.; Melkemi, N.; Salah, T.; Ghalem, S. Combined QSAR, molecular docking and molecular dynamics study on new Acetylcholinesterase and Butyrylcholinesterase inhibitors. **2018**, doi:10.1016/j.compbiolchem.2018.03.021.

20. Bitencourt-Ferreira, G.; de Azevedo, W.F. Molegro virtual docker for docking. In *Methods in Molecular Biology*; Humana Press Inc., 2019; Vol. 2053, pp. 149–167.

21. Aps, M. *molegro data modeller user manual MDM 2009.2.0 for Windows, Linux, and Mac OS X*; 2007;

22. Sadeghi, F.; Afkhami, A.; Madrakian, T.; Ghavami, R. Computational study to select the capable anthracycline derivatives through an overview of drug structure-specificity and cancer cell line-specificity. *Chem. Pap.* **2021**, *75*, 523–538, doi:10.1007/s11696-020-01321-z.

23. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Genet.* **2006**, *65*, 712–725, doi:10.1002/prot.21123.

24. Swiss PDB Viewer - Home Available online: https://spdbv.vital-it.ch/ (accessed on May 25, 2021).

25. S, B.; F, B.; Z, D. Docking studies of (-)-Epigallocatechin-3-gallate: A Potential Non-Competitive Pancreatic Lipase Inhibitor. *Res. J. Pharm. Biol. Chem. Sci.* **2016**, *7*, 2493–2505.

26. Singh, S.P.; Ranjan Deb, C.; Ahmed, S.U.; Saratchandra, Y.; Kumar Konwar, B. Molecular docking simulation analysis of the interaction of dietary flavonols with heat shock protein 90. *J. Biomed. Res.* **2016**, *30*, 67–74, doi:10.7555/JBR.30.20130158.

27. pkCSM Available online: http://biosig.unimelb.edu.au/pkcsm/prediction (accessed on Jun 7, 2021).

28. SwissADME Available online: http://www.swissadme.ch/ (accessed on Jun 7, 2021).

29. Molinspiration Cheminformatics Available online: https://www.molinspiration.com/ (accessed on Jun 7, 2021).

30. ERRAT - help - DOE-MBI Structure Lab UCLA Available online: https://servicesn.mbi.ucla.edu/ERRAT/help (accessed on Jun 7, 2021).

31. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291, doi:10.1107/s0021889892009944.

32. PDBsum entry: 6m3m Available online: http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl (accessed on Jun 7, 2021).

33. Thomsen, R.; Christensen, M.H. MolDock: A New Technique for High-Accuracy Molecular Docking. **2006**, doi:10.1021/jm051197e.

34. Copeland, R.A. Conformational adaptation in drug-target interactions and residence time. *Future Med. Chem.* 2011, *3*, 1491–1501.

35. Guan, L.; Yang, H.; Cai, Y.; Sun, L.; Di, P.; Li, W.; Liu, G.; Tang, Y. ADMET-score-a comprehensive scoring function for evaluation of chemical drug-likeness. *Medchemcomm* **2019**, *10*, 148–157, doi:10.1039/C8MD00472B.

Figures

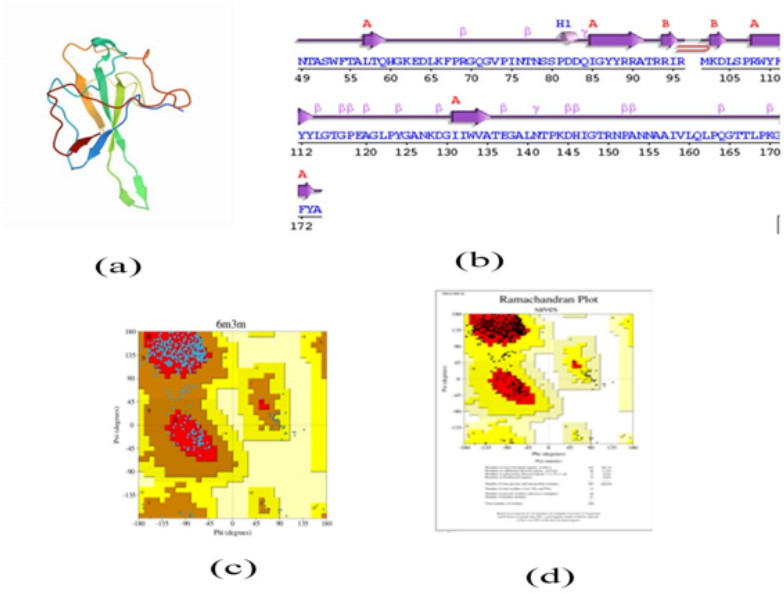

Figure 1

Structural information of N protein (a) cartoon 3d structure of N protein (b) secondary structure of 3pty mapped obtained using PDBsum (c) and (d) Ramachandran Plot

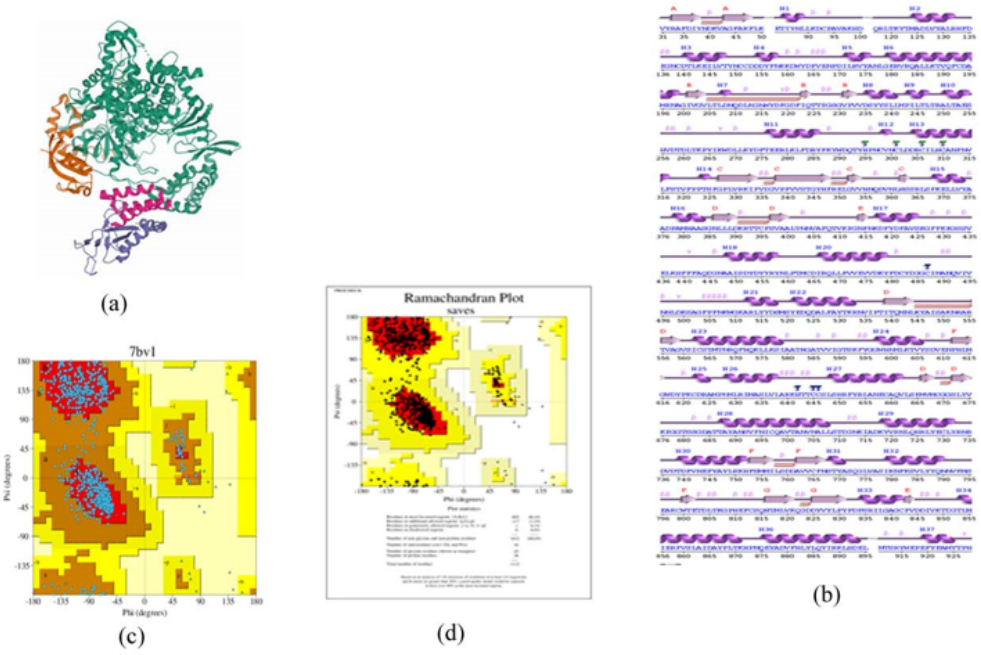

Figure 2

Structural information of RdRp protein (a) cartoon 3d structure of RdRp protein (b) secondary structure of 3pty mapped obtained using PDBsum (c) and (d) Ramachandran Plot

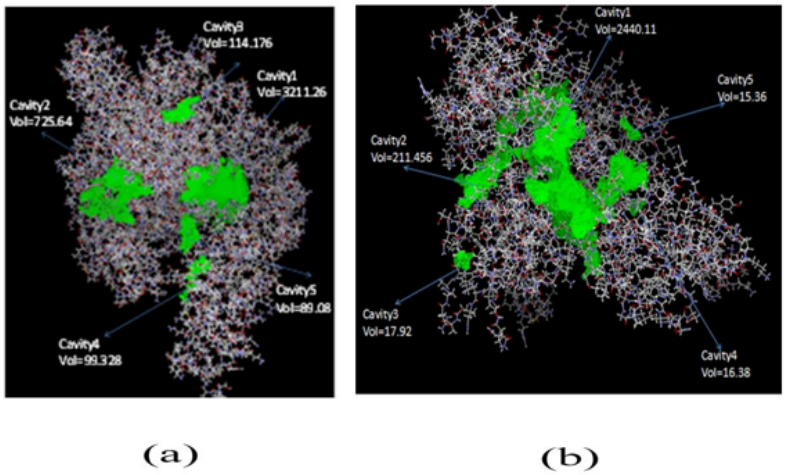

Figure 3

Active sites of Proteins (a) RdRp (b) Nucleocapsid

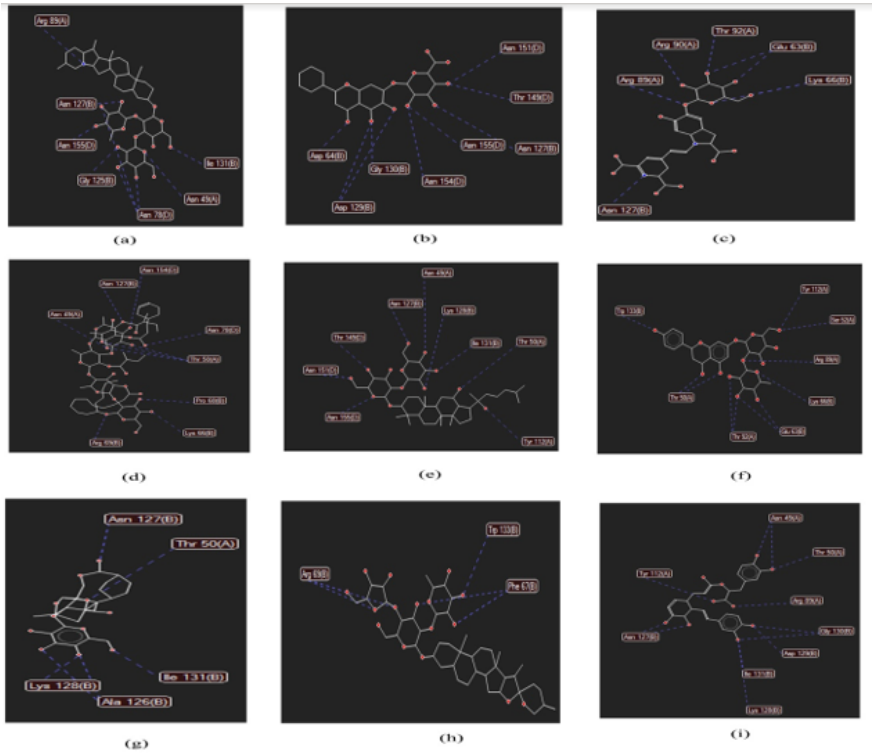

Figure 4

Hydrogen Bond interactions of ligands with N protein (a) Alpha solanine (b) Baicalin (c) Betanin (d) Cairicoside I (e) Ginsenoside rb1 (f) Naringin (g) Paeoniflorin (h) Polyphyllin I (i) Salvianolic acid A

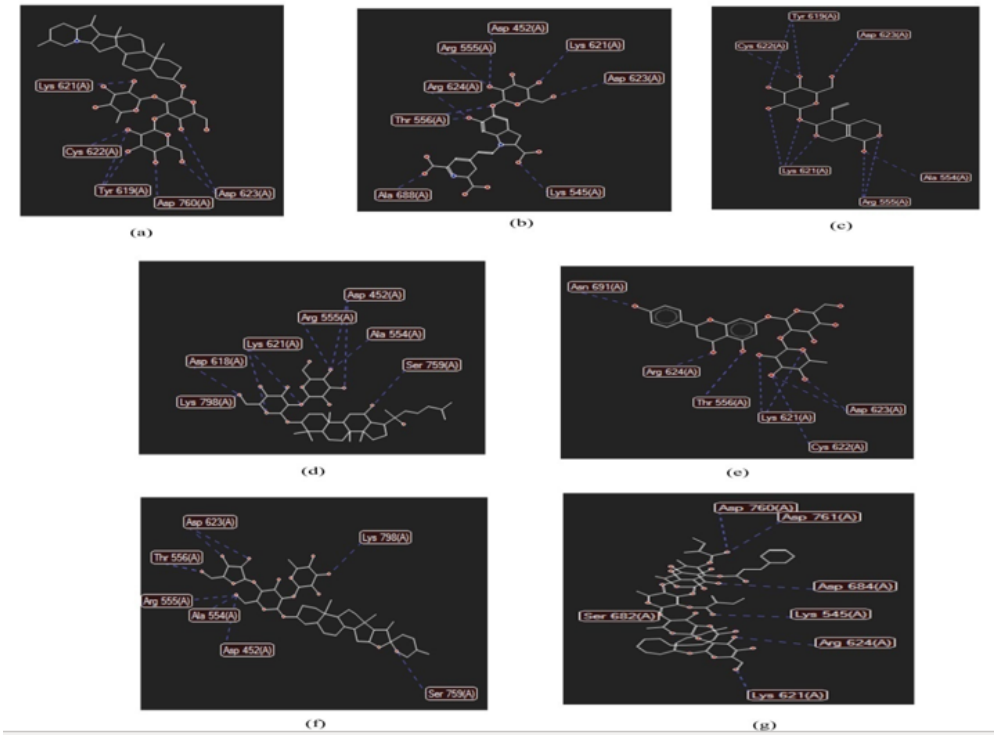

Figure 5

Hydrogen bond interactions of ligands with RdRp protein (a) Alpha solanine (b) Betanin (c) Gentiopicroside (d) Ginsenoside rb1 (e) Naringin (f) Polyphyllin I (g) Cairicoside I

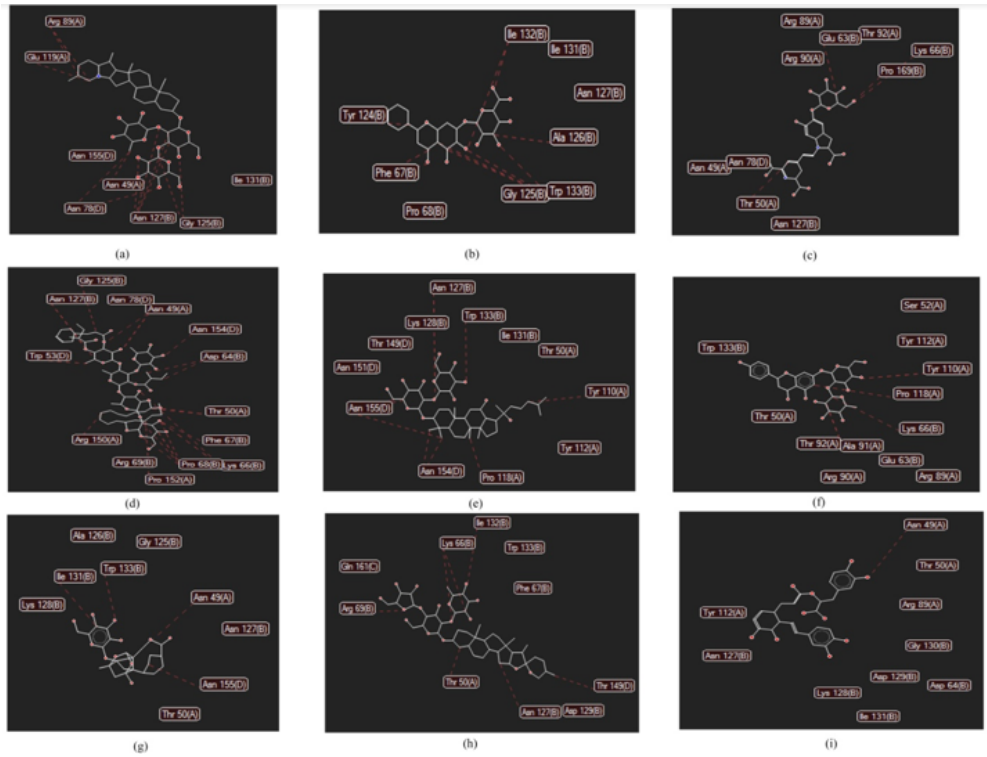

Figure 6

Steric interactions of ligands with N protein (a) Alpha solanine (b) Baicalin (c) Betanin (d) Cairicoside I (e) Ginsenoside rb1 (f) Naringin (g) Paeoniflorin (h) Polyphyllin I (i) Salvianolic acid A

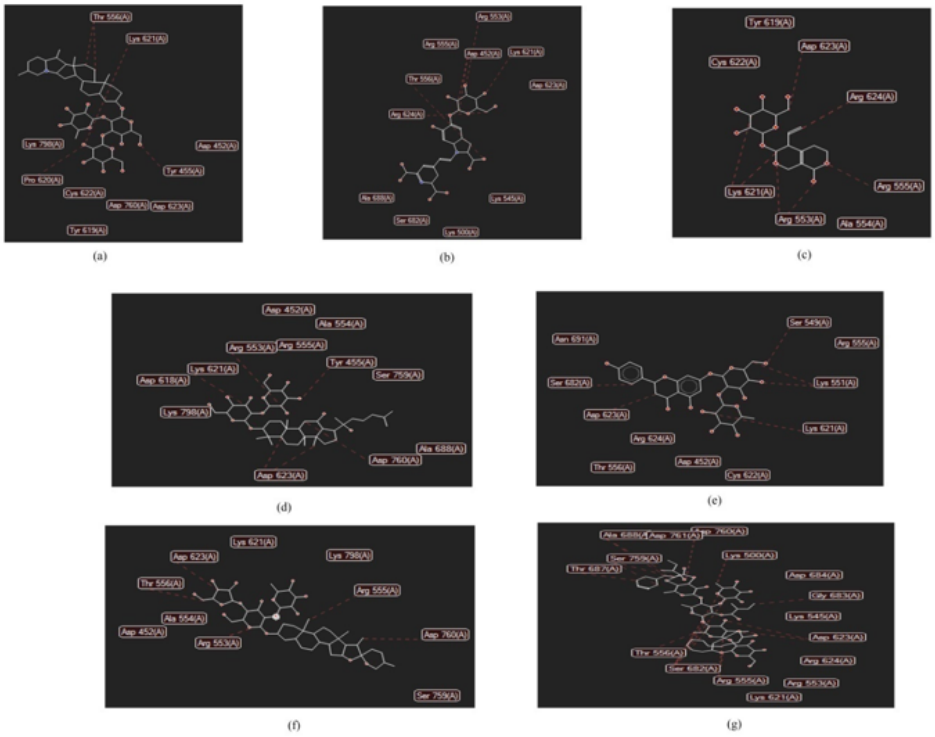

Figure 7

Steric interactions of ligands with RdRp protein (a) Alpha solanine (b) Betanin (c) Gentiopicroside (d) Ginsenoside rb1 (e) Naringin (f) Polyphyllin I (g) Cairicoside I

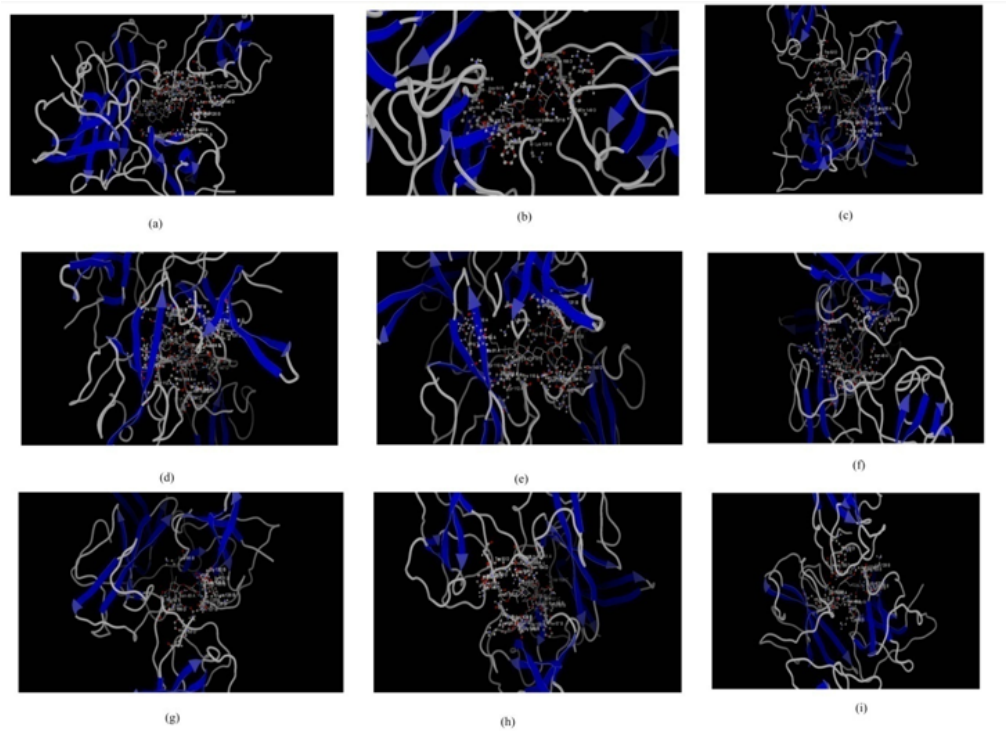

Figure 8

Ligand docked against the crystal protein structure of COVID-19 Nucleocapsid (a) Alpha solanine (b) Baicalin (c) Betanin (d) Cairicoside I (e) Ginsenoside rb1 (f) Naringin (g)Paeoniflorin (h) Polyphyllin I (i) Salvianolic acid A

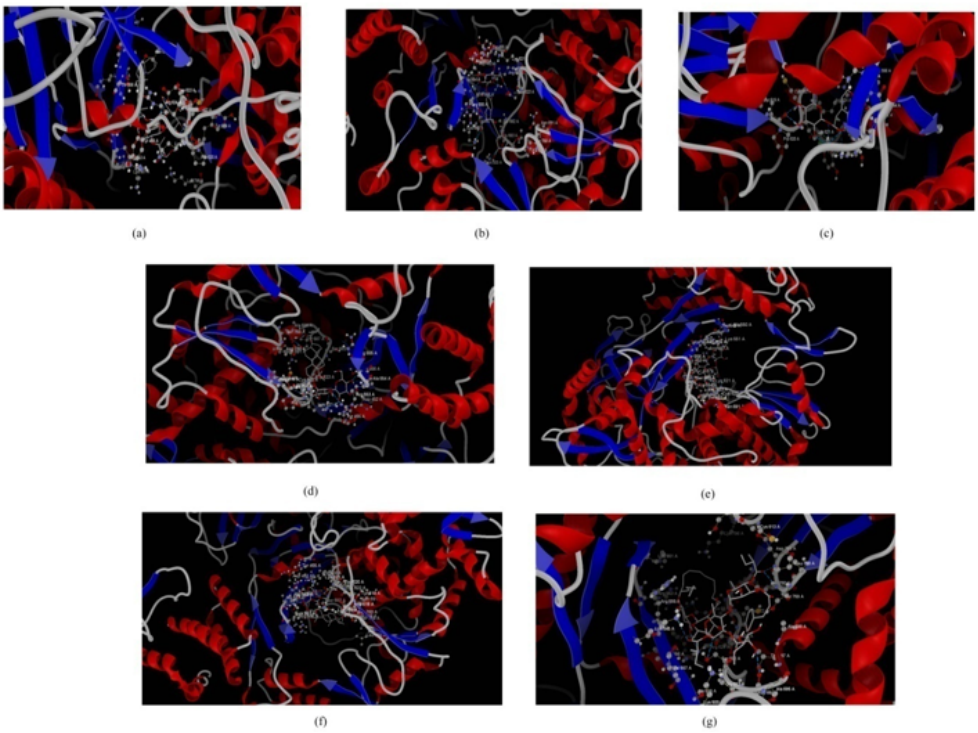

Figure 9

Ligand docked against the crystal protein structure of COVID-19 RdRp (a) Alpha solanine (b) Betanin (c) Gentiopicroside (d) Ginsenoside rb1 (e) Naringin (f) Polyphyllin I (g)Cairicosidel

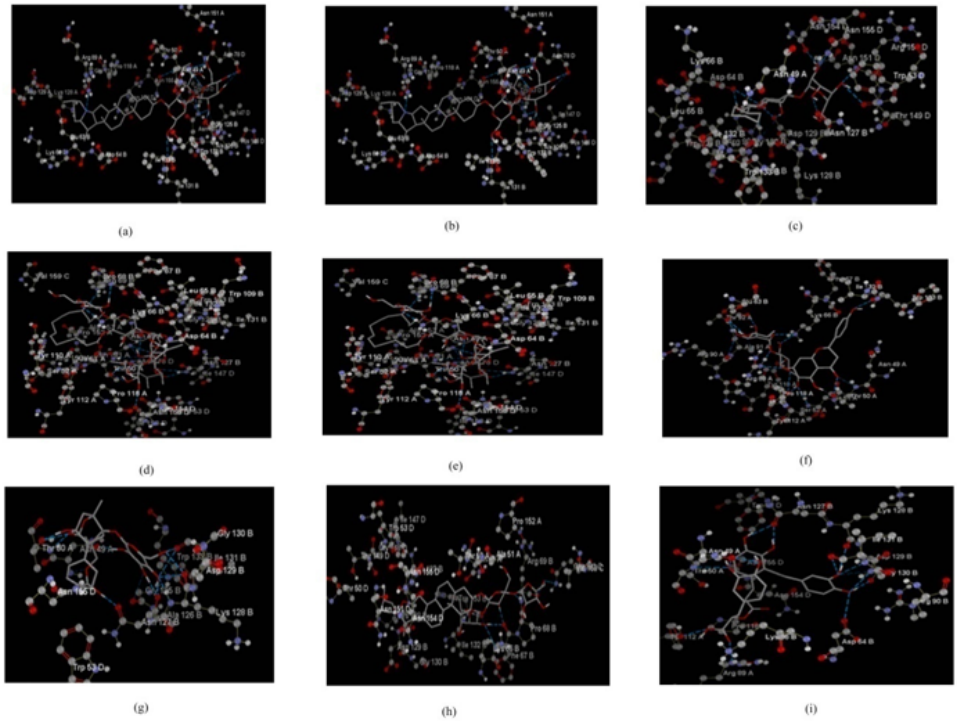

Figure 10

The amino acid residues around active site and docked against N protein (a) Alpha solanine (b) Baicalin (c) Betanin (d) Cairicoside I (e) Ginsenoside rb1 (f) Naringin (g)Paeoniflorin (h) Polyphyllin I (i) Salvianolic acid A

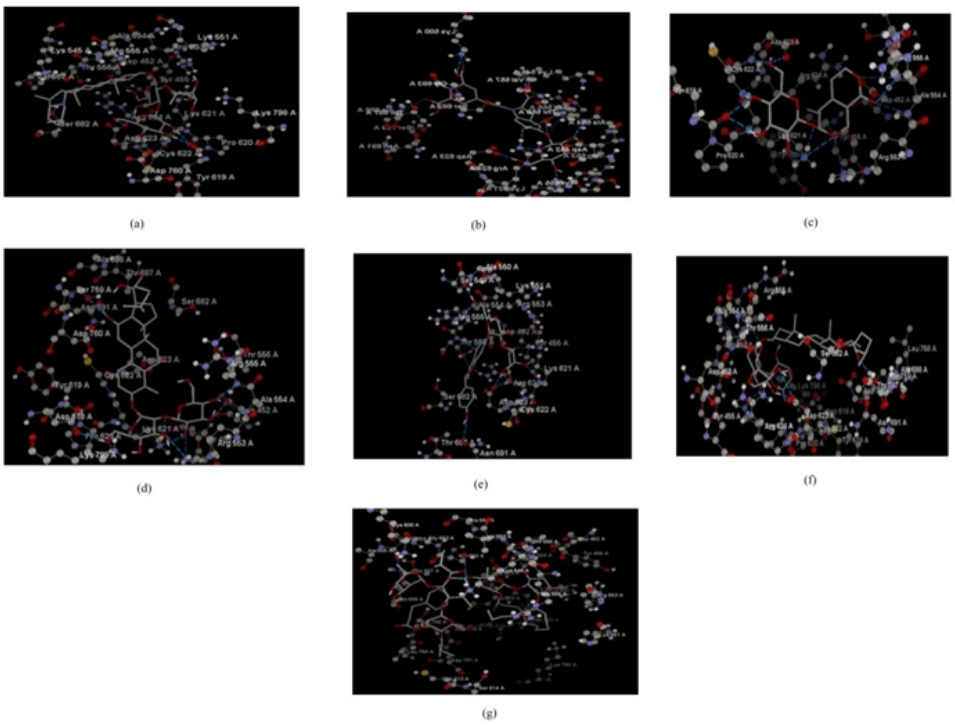

Figure 11

The amino acid residues around active site and docked against RdRp protein (a) Alpha solanine (b) Betanin (c) Gentiopicroside (d) Ginsenoside rb1 (e) Naringin (f) Polyphyllin I (g)Cairicosidel

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx